

cdNA sequence:

```

1  ATGAGCTTAA AGTGTCTCTG TCTTGCTTGC AGGCTACAAC CCATTTGCCC
51 CATTGAAGGT CGACTGGGTG GAGCCCGCAC TCAGGCTGAA TTCCCACTTC
101 GCGCCCTGCA GTTTAAGCGT GGCCTGCTGC ACGAGTTCCG GAAGGGCAAC
151 GCTTCCAAGG AGCAGGTTCTG CCTCCATGAC CTGGTCCAGC AGCTCCCCAA
201 GGCCATTATC ATTGGGGTGA GGAAAGGAGG CACAAGGGCC CTGCTTGAAA
251 TGCTGAACCT ACATCCGGCA GTAGTCAAAG CCTCTCAAGA AATCCACTTT
301 TTTGATAATG ATGAGAATTA TGGTAAGGGC ATTGAGTGGT ATAGGAAAAA
351 GATGCCTTTT TCCTACCTC AGCAAATCAC AATTGAAAAG AGCCCAGCAT
401 ATTTTATCAC AGAGGAGGTT CCAGAAAGGA TTTACAAAAT GAACTCATCC
451 ATCAAGTTGT TGATCATTGT CAGGGAGCCA ACCACAAGAG CTATTTCTGA
501 TTATACTCAG GTGCTAGAGG GGAAGGAGAG GAAGAACAAA ACTTATTACA
551 AGTTTGAGAA GCTGGCCATA GACCCTAATA CATGCGAAGT GAACACAAAA
601 TACAAAGCAG TAAGAACCAG CATCTACACC AAACATCTGG AAAGGTGGTT
651 GAAATACTTT CCAATTGAGC AATTTTCATG CGTCGATGGA GATCGCCTCA
701 TCACGGAACC TCTGCCAGAA CTTTCAGCTG TGGAGAAGTT CCTAAATCTG
751 CCTCCAAGGA TAAGTCAATA CAATTTATAC TTCAATGCTA CCAGAGGGTT
801 TTAAGTCTTG CGGTTTAATA TTATCTTTAA TAAGTGCCTG GCGGGCAGCA
851 AGGGGCGCAT TCATCCAGAG GTGGACCCCT CTGTCATTAC TAAATTGCGC
901 AAATCTTTT ATCCTTTTAA TCAAAAATTT TACCAGATCA CTGGGAGGAC
951 ATTGAAGTGG CCCTAAGGGC (SEQ ID NO:1)

```

FEATURES:

5'UTR:

Start Codon: 1

Stop Codon: 964

3'UTR: 967

HOMOLOGOUS PROTEIN:

Top 10 BLAST Hits:

gi 4826764 ref NP_005105.1	heparan sulfate (glucosamine) 3-O-s...	303	1e-81
gi 6754246 ref NP_034604.1	heparan sulfate (glucosamine) 3-O-s...	297	1e-79
gi 9957244 gb AAG09283.1	(AF177430) 3-O-sulfotransferase [Ratt...	293	3e-78
gi 7293568 gb AAF48941.1	(AE003511) CG7890 gene product [Droso...	251	7e-66
gi 5174463 ref NP_006034.1	heparan sulfate (glucosamine) 3-O-s...	248	6e-65
gi 5174467 ref NP_006032.1	heparan sulfate (glucosamine) 3-O-s...	247	1e-64
gi 5174465 ref NP_006033.1	heparan sulfate (glucosamine) 3-O-s...	247	1e-64
gi 9055264 ref NP_061275.1	D-glycosaminyl 3-O-sulfotransferase...	240	2e-62
gi 4835727 gb AAD30210.1	AF105378_1 (AF105378) heparan sulfate ...	235	6e-61
gi 7503118 pir T33493	hypothetical protein F40H3.5 - Caenorhab...	187	2e-46

EST:

gi 6990086	/dataset=dbest /taxon=960...	672	0.0
gi 5362532	/dataset=dbest /taxon=9606 ...	660	0.0
gi 4682357	/dataset=dbest /taxon=9606 ...	571	e-161

FIGURE 1A

EXPRESSION INFORMATION FOR MODULATORY USE:

gi|6990086 / lung

gi|5362532 / lung carcinoid

gi|4682357 /lung carcinoid

Tissue expression

Human Brain

Human bone marrow

Human colon

Human fetal brain

Human fetal heart

Human fetal liver

Human fetal lung

Human pancreas

Human placenta

FIGURE 1B

```

1 MSLKCLCLAC RLQPICPIEG RLGGARTQAE FPLRALQFKR GLLHEFRKGN
51 ASKEQVRLHD LVQQLPKAII IGVRKGGTRA LLEMLNLHPA VKASQEIHF
101 FDNDENYGKG IEWYRKKMPF SYPQQITIEK SPAYFITEEV PERIYKMNSS
151 IKLLIIVREP TTRAISDYTQ VLEGKERKNK TYYKFEKLAI DPNTCEVNTK
201 YKAVRTSIYT KHLERWLKYF PIEQFHVDG DRLITEPLPE LQLVEKFLNL
251 PPRISQYNLY FNATRGFYCL RFNIIFNKCL AGSKGRIHPE VDPSVITKLR
301 KFFHPPFNQKF YQITGRTLNW P (SEQ ID NO:2)

```

FEATURES:

Functional domains and key regions:

Prosite results:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site

Number of matches: 4

1	50-53	NASK
2	148-151	NSSI
3	179-182	NKTY
4	262-265	NATR

[2] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 4

1	2-4	SLK
2	150-152	SIK
3	161-163	TTR
4	314-316	TGR

[3] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

27-30 TQAE

[4] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 3

1	23-28	GGARTQ
2	72-77	GVRKGG
3	76-81	GGTRAL

FIGURE 2A

BLAST Alignment to Top Hit:

Score = 774 (277.5 bits), Expect = 6.0e-76, P = 6.0e-76

Identities = 137/258 (53%), Positives = 189/258 (73%), Frame = +2

>TRM|O14792|O14792 /def="HEPARAN SULFATE 3-O-SULFOTRANSFERASE-1 PRECURSOR. "
 /org=Homo_sapiens /date=01-MAY-2000 /mol_type=PRT /len=307
 /gene_name=3OST /mol_weight=35773 /seq_status=PRELIMINARY
 Length = 307

Query: 191 QQLPKAIIIGVRKGGTRALLEMLNLHPAVKASQEIHFDDNDENYGKGIWYRKKMPFSY 370
 QQLP+ IIIGVRKGGTRALLEML+LHP V A E+HFFD +E+Y G+ WY +MPFS+
 Sbjct: 52 QQLPQTIIIGVRKGGTRALLEMLSLHPDVAANAENEVHFFDWEEHYSHGLGWYLSQMPFSW 111

Query: 371 PQQITIEKSPAYFITEEVPERIYKMNSSIKLLIIVREPTTRAISDYTQVLEGKERKNKTY 550
 P Q+T+EK+PAYF + +VPER+Y MN SI+LL+I+R+P+ R +SDYTQV +K+K Y
 Sbjct: 112 PHQLTVEKTPAYFTSPKVPERVYSMNPSIRLLLLILRDPSEVLSDYTQVFYNHMQKHKPY 171

Query: 551 YKFEKLAIDPNTCEVNTKYKAVRTSIYTKHLERWLKYFPIEQFHVDGDRLITEPLPELQ 730
 E+ + +N YKA+ S+Y H++ WL++FP+ H+VDGDRLI +P PE+Q
 Sbjct: 172 PSIEEFLVRDG--RLNVDYKALNRSLYHVMQNWLRFFPLRHHIVDGDRLIRDPFPEIQ 229

Query: 731 LVEKFLNLPPRISQYNLYFNATRGFYCLRFNIIFNKCLAGSKGRIHPEVDPSVITKLRKF 910
 VE+FL L P+I+ N YFN T+GFYCLR + ++CL SKGR HP+VDP ++ KL ++
 Sbjct: 230 KVERFLKLSPQINASNIFYNKTGKGFYCLR-DSGRDRCLHESKGRAHPQVDPKLLNLKHEY 288

Query: 911 FHPFNQKFYQITGRTLNW 964
 FH N+KF+++ GRT +W
 Sbjct: 289 FHEPNKKFFELVGRTFDW 306 (SEQ ID NO:4)

Hmmer search results (Pfam):

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
PF02004	Protein of unknown function	4.3	6.5	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF02004	1/1	133	147	..	266 280 .]	4.3	6.5

FIGURE 2B

1	ATTAGCTTCC	AATCATTTAC	CTTTTACTTA	GTAATTGATC	TAATGATCAC
51	TAATGCATTA	TTATTTAGTT	GATGATTCTT	TTCATTTTTT	TAACTCTGTC
101	TCTAGTCTCT	AAGGGGATAG	CTTTTATTTG	GAATTGAATT	GTTTGGTGGG
151	CTTTCTAAAA	GCCTCTCACT	TCAGACTTTG	AGATTATGTC	TGAAGGTAAC
201	AGGCTTATTT	AGGCCCACTC	TCCAGTAACT	GAAGACCCTG	CTTTCTGGGA
251	GGGAGACAGA	GGTTACTTCT	ACCATCCCTT	CCAATCCTAA	ACCTGTATGA
301	TTTTTCAGTC	TGGGACCCAT	ACTCAGAATC	CATGCTTTCA	GAAGTGGGAA
351	AGAATATGAT	ATTTTCTCAA	ATTTTCACAT	TCTATCTTGA	GTTAGGGAGT
401	CCAAAAAGCG	ACTATTCTGC	AGGATGTGAT	CTCCCAGGGT	AGAAGATAGA
451	AAGAGGAAGG	AAGTAAAGAA	GGAAAATGAC	CCTTTCTACA	AGTGGGGAAA
501	TTCCATTTGA	CCTCAAACAA	AGCAGAGACT	GTCTATATCA	GCCACTCTCA
551	GCCAGGGTAC	TATGAAAGAA	TTAAATCCTA	CAAAAAAGAA	TTTGAGTGAC
601	TGTTTCCTCA	ATTCTTCCAA	GGATGGTACT	AGCATCATTC	TAGGTGCTTA
651	GGACAGAAAT	CCATCAATGG	ATGCCTTATG	GAATTAGAGC	TTAATTCTCA
701	ACCAGAACCC	AAGAAGAACT	GAAAGATGAA	CTTGTATTAT	TCCAATCAGT
751	GTCACAATTA	AAAGCATCTT	TGCCTATGTA	TCTATTGATA	ATTTTACATC
801	CTCCATTTAA	AGCCCTAGTA	CATTAATCTC	ATTAACAAAT	TTATAAAAAAC
851	AAAATTTCATG	TTTCTCTAAA	CTATTAACCG	GGTTAAATCC	TGTTTTTTTAA
901	AAGCTGTCTA	GGCCAGGCAC	AGTAGCTCAC	GCCTGTAATC	CCAGCACTTT
951	GGGAGGCTGA	GGCAGGCGAA	TCACGAGATC	AGGAGTTCAA	GACCAGCCAG
1001	GCCAACATGG	TGAAACCTTG	TCTCTACTAA	AAATACAAAA	ATTAGCTGGG
1051	TATGGTGGCG	CAGGCCTGTA	ATCCCAGCTA	CTCGGGAGGC	TGAGGCAGGA
1101	GAATCTCTTG	AACCCAGGAG	ACAGAGATTG	CAGTGAGCCA	AGATCGTGCC
1151	ACTGCACTGC	AGCCTAGGCA	ACAGACCAAG	ACTCCGTCTC	AAAAAAAAAAA
1201	GAATAAAAAAG	TTGTCTATAT	TTTCACACTT	TCCACAATGA	GCATGAGTTG
1251	TTTTAAAAAT	CATAAAAAAG	AAACATCGTG	AAAAGTAGTA	TACATTGATA
1301	TTTTTCCTTA	AGCATTATGA	TAGATAGCTG	TTTAAACAGA	ACAAAGACCA
1351	AGACCATGCT	CCTCAATTCT	GCAGAACAGG	CTGAGTGTAT	TAGTCCGTTT
1401	TCACAGTGCT	ATAAAGACAT	ACCTGAGACT	GAGTAATTTA	TAAAGAAAAA
1451	AGGTTTAAAT	GACACACAGT	TCTGCATGGC	TGGGGAAGCC	TCAGAAAACT
1501	TACAATCATG	GCAGAAGGCA	AAGAAGAAGC	AAGGCACGTC	TTACTTGGTG
1551	GCAGGAGAGA	GAGGGAGCTT	GCAGGGGGCG	GTGCCACACA	GTTTTAAACC
1601	ATCAAATCTC	ATGAGAACTC	ACTATCATGA	AAACAAGGGG	TAAATACACC
1651	CCCATAATCC	AGTCACCTCC	CACCAAGCCC	CTCCTCCGAC	ATGTGGGGAT
1701	TACAATTCGG	GATGAGATTT	GGGTGGGGGC	ACAGAGCCAA	ACCATATCAC
1751	TGGGCATGAC	CTTGAGGTTG	TTTCTCATCT	CAGAAAACAA	GAAAGATGCA
1801	ATACAGTCTC	TTGGGAAAAG	CAAGCAACAG	CCTCATTGCC	ACAGAGGGGG
1851	AGACACAGAT	TCCAAATTAT	TAGAATAACT	GGAAGCTTTC	AAGTGTAAGA
1901	ATTGGTTTAA	CAGCCTTTTT	GACTGATATT	ATTTAATTTT	ACCAAGAAGG
1951	CTAAAATGCC	CTCACAGATC	AACTTAGGGG	AATTATAATG	AACTTCAGTT
2001	CAATTCAGAC	TATACCTAAA	AGGAAACTCA	ATTTGCTAAC	CATATATGTT
2051	AGCCATGACA	AATTA AACAG	TCACCATCGT	CTACTATCAT	TGTGACTGTT
2101	ACCACATCTT	TCTCCCTGAG	AAAAGCAGAG	ATGGTTGTTC	ACTATTCAGG
2151	ATAATACTGA	AGTGGAATC	CTCCTGTCTG	GCTATATCCA	TTGCACTCCT
2201	TCCTTAATGA	GATTGAGTTC	CTGATTTTAA	TGGGCTTGGC	AATGAGGGCT
2251	TGAGGTTTCT	GGCCCTGTCA	AGGTCTTGTT	GATGCCTGGT	CCCAGGTGTG
2301	GTAGGTGATA	TACAGCACTT	GCTGATGGCA	ATTGGGTTTG	ATTCTATATT
2351	CAGCAAAGTG	GATATATAAT	CCTGACCTCT	TTAGATAGAA	AGAGAAAGAG
2401	AGGCAGAAGA	AATATAGTAT	TCTTCTGGCT	ATCCTCAAGG	CCCAGGGCAG
2451	AGAGTCTCAG	AATGAAAATC	TCAGCAAGTT	CCAAGATTGG	AATTTTGCAG
2501	GTTGATGATG	CAACAGCCCC	GGGGCAGAAA	CTGGGACCTC	CTTTCAGATT
2551	ATATCTCAAA	GATTTTCAAG	AGCCATCTGA	GTGCTGCCGA	GCTGCAAGAA
2601	AATAATACCA	CACAAAATGT	GAAACACATG	GCCTCCCTGC	TACCCTTCCA
2651	CCTCCAGCT	GAAGATTATA	ATCTCCTGCC	TTTCACTTTT	TCTTAATGAT

FIGURE 3A

```

2701 TTTAACTGGT GAGCTGTAA AAAGCTATTA GTATGGCTGG TGCCACTTGT
2751 CTATCCTGTA CTGCAAACAG AAGTGCACGC CGTAGTCAAT TAAGTGCTTG
2801 GAGAATAAAA AATTTTAAGG AGCACTAATA AAAAAATTCA TCAATTATGT
2851 GTGCTCCATT TAATACATGG TTGCTTAAAA TAAATTTTCC CAAACATATG
2901 TTCATTATGG ATTGCAGCAG GCTGGGAACC AGTGGCTTTA TTTATGCATT
2951 TAAAGTCTTG GTCTGACTGG GGAACCAGAA AAATGAAAAG TTAGTTGCAA
3001 TGAGCTTAAA GTGTCTCTGT CTTGCTTGCA GGCTACAACC CATTTGCCCC
3051 ATTGAAGGTC GACTGGGTGG AGCCCGCACT CAGGCTGAAT TCCCACTTCG
3101 CGCCCTGCAG TTTAAGCGTG GCCTGCTGCA CGAGTTCGGG AAGGGCAACG
3151 CTTCCAAGGA GCAGGTTTCGC CTCCATGACC TGGTCCAGCA GCTCCCCAAG
3201 GCCATTATCA TTGGGGTGAG GAAAGGAGGC ACAAGGGCCC TGCTTGAAAT
3251 GCTGAACCTA CATCCGGCAG TAGTCAAAGC CTCTCAAGAA ATCCACTTTT
3301 TTGATAATGA TGAGAATTAT GGTAAAGGGCA TTGAGTGGTA TAGGAAAAAG
3351 ATGCCTTTTT CCTACCCTCA GCAAATCACA ATTGAAAAGA GCCCAGCATA
3401 TTTTATCACA GAGGAGGTTC CAGAAAGGAT TTACAAAATG AACTCATCCA
3451 TCAAGTTGTT GATCATTGTC AGGGAGCCAA CCACAAGAGC TATTTCTGAT
3501 TATACTCAGG TGCTAGAGGG GAAGGAGAGG AAGAACAAAA CTTATTACAA
3551 GTTTGAGAAG CTGGCCATAG ACCCTAATAC ATGCGAAGTG AACACAAAAT
3601 ACAAAGCAGT AAGAACCAGC ATCTACACCA AACATCTGGA AAGGTGGTTG
3651 AAATACTTTC CAATTGAGCA ATTTTCATGTC GTCGATGGAG ATCGCCTCAT
3701 CACGGAACCT CTGCCAGAAC TTCAGCTCGT GGAGAAGTTC CTAAATCTGC
3751 CTCCAAGGAT AAGTCAATAC AATTTATACT TCAATGCTAC CAGAGGGTTT
3801 TACTGCTTGC GGTTTAATAT TATCTTTAAT AAGTGCCTGG CGGGCAGCAA
3851 GGGGCGCATT CATCCAGAGG TGGACCCCTC TGTCACTACT AAATTGCGCA
3901 AATTCTTTCA TCCTTTTAAT CAAAAATTTT ACCAGATCAC TGGGAGGACA
3951 TTGAAGTGGC CCTAAAATAA TATGTCATAC AACACTATGT GTTGTGCCTG
4001 GAGACACACA ATGTCTCCTG TAGATTAATA TATGCACTTT TCCTAGGCAG
4051 AGCTATCCAA GTCATTTTTC CATGTATATT TGTACATACG CAGTGTGTGA
4101 CCAAATATAA GATCAGTTCT TTTTCTACTG AAAATTTACG AAAAAAAAAA
4151 AATTGCTGTC TGCATAGTCG CATCTTTTAA GCTATTTACA AAAGAGAAGA
4201 GGTGGTGGTA TTGGGGGAAA GTGACTTCAG CTATTCTCAA AGAGTTAGTC
4251 TTCCTTTGAT TCAGAATTG TACCCGCCCA TTTTCATAGA TTTAAGCCAA
4301 AAGATAAATG TGTGAAAATG TACCAATGGC TGCGAAGCTT CAGGAAGTAG
4351 AGGATCCAGT GATGCATTTT TTTTTCCTA AGGGAAAGCT GGCTCTTTAA
4401 TTCAGATGCT GAATTGGTGC CATGAAAACA GAAAATGCTA TTTTCTTATT
4451 ATTTAAAAGA ACGTCTTATC TCATAAAATT GACATTGTTC CAAAGTTCTT
4501 GTGGTGATTT TGCATATTG TTTTCTCGTA TGGACCATGG TGCACTTGT
4551 AGCATGTCAA TCACACATTG GAAAGTCAAG TCCTTTTACT TCCATGTTGT
4601 ATGTCAACAG AGAGAAATGT CATGTACATA ATGTATATTG TTGTAAATAC
4651 TGGTTTCACA CTAAGTAATT CTATTTTGTA AACTGAATAT GGCTATTTAA
4701 TTTATTGTGA AAATTAATTT TATTGTGGTA TTTAAAAATG GAATGGATTA
4751 AAATTACTCT ATGTGCAATT TTTTTTTTTT TTAATCATTG TGTTTTACGT
4801 GCCCCCTGCT GGCTTCCAAA ATGGAAGCTG TTTACGTGCA TATGAGAGCA
4851 CTTGGAAAAGA TGTGCTTCCC TGCTGGATTT CTGTACCCCA GTGAAAATGT
4901 ATTTATGAAG TGAGGTTGAG TATATTAATA AAGAAAAACC TCAACCATCT
4951 GGAAATCAAG TATAATAGCC ACCTCAAAGA ACCCTAGTGC TGCTCTGCTA
5001 CAACTTTGTA ACAATTAATT TACTCGCAGT TGCTGCTGCT CAGG (SEQ ID NO:3)

```

FEATURES:

Start: 3000
Exon: 3000-3962
Stop: 3963

FIGURE 3B

SNPs:

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
155	C	G	Beyond ORF(5')			
370	A	T	Beyond ORF(5')			
2775	G	A	Beyond ORF(5')			
4240	A	G	Beyond ORF(3')			

Context:

DNA
Position

155	ATTAGCTTCCAATCATTTACCTTTTACTTAGTAATTGATCTAATGATCACTAATGCATTA TTATTTAGTTGATGATTCTTTTCATTTTTTAACTCTGTCTCTAGTCTCTAAGGGGATAG CTTTTATTTGGAATTGAATTGTTTGGTGGGCTTT [C,G] TAAAAGCCTCTCACTTCAGACTTTGAGATTATGTCTGAAGGTAACAGGCTTATTTAGGCC CACTCTCCAGTAACTGAAGACCCTGCTTTCTGGGAGGGAGACAGAGGTTACTTCTACCAT CCCTTCCAATCCTAAACCTGTATGATTTTTCTGCTGGGACCCATACTCAGAATCCATGC TTTCAGAAGTGGGAAAGAATATGATATTTTCTCAAATTTTCACATTCTATCTTGAGTTAG GGAGTCCAAAAGCGACTATTCTGCAGGATGTGATCTCCAGGGTAGAAGATAGAAAGAG
370	TGATGATTCTTTTCATTTTTTTAACTCTGTCTCTAGTCTCTAAGGGGATAGCTTTTATTT GGAATTGAATTGTTTGGTGGGCTTTCTAAAAGCCTCTCACTTCAGACTTTGAGATTATGT CTGAAGGTAACAGGCTTATTTAGGCCCACTCTCCAGTAACTGAAGACCCTGCTTTCTGGG AGGGAGACAGAGGTTACTTCTACCATCCCTTCCAATCCTAAACCTGTATGATTTTTCTAGT CTGGGACCCATACTCAGAATCCATGCTTTCAGAAGTGGGAAAGAATATGATATTTTCTCA [A,T] ATTTTCACATTCTATCTTGAGTTAGGGAGTCCAAAAGCGACTATTCTGCAGGATGTGAT CTCCCAGGGTAGAAGATAGAAAGAGGAAGGAAGTAAAGAAGGAAAATGACCCTTTCTACA AGTGGGGAAATTCCATTTGACCTCAAACAAAGCAGAGACTGTCTATATCAGCCACTCTCA GCCAGGGTACTATGAAAGAATTAAATCCTACAAAAAAGAATTTGAGTGACTGTTTCCTCA ATTCCTCAAGGATGGTACTAGCATCATTCTAGGTGCTTAGGACAGAAATCCATCAATGG
2775	CAAGTTCCAAGATTGGAATTTTGCAGGTTGATGATGCAAACAGCCCGGGGCAGAACTGG GACCTCCTTTTCAGATTATATCTCAAAGATTTTCAAGAGCCATCTGAGTGCTGCCGAGCTG CAAGAAAATAATACCACACAAAATGTGAAACACATGGCCTCCCTGCTACCCTTCCACCTC CCAGCTGAAGATTATAATCTCCTGCCTTTCACTTTTTCTTAATGATTTTAACTGGTGAGC TGTTAAAAAGCTATTAGTATGGCTGGTGCCACTTGTCTATCCTGTACTGCAAACAGAAGT [G,A] CACGCCGTAGTCAATTAAGTGCTTGGAGAATAAAAAATTTTAAGGAGCACTAATAAAAAA ATTCATCAATTATGTGTGCTCCATTTAATACATGGTTGCTTAAAATAAAATTTCCCAAAC ATATGTTTATTATGGATTGCAGCAGGCTGGGAACCACTGGCTTTATTTATGCATTTAAAG TCTTGGTCTGACTGGGGAACCAAGAAAATGAAAAGTTAGTTGCAATGAGCTTAAAGTGTG TCTGTCTTGCTTGCAGGCTACAACCCATTTGCCCCATTGAAGGTCGACTGGGTGGAGCCC
4240	CTGGGAGGACATTGAACTGGCCCTAAAATAATATGTCATACAACACTATGTGTTGTGCCT GGAGACACACAATGTCTCCTGTAGATTAAAATATGCACTTTTCTAGGCAGAGCTATCCA AGTCATTTTTCCATGTATATTTGTACATACGCAGTGTGTGACCAAATATAAGATCAGTTC

FIGURE 3C

TTTTCTACTGAAAATTTACGAAAAAAAAAAATTGCTGTCTGCATAGTCGCATCTTTTA
AGCTATTTACAAAAGAGAAGAGGTGGTGGTATTGGGGGAAAGTGACTTCAGCTATTCTCA
[A,G]
AGAGTTAGTCTTCCTTTGATTCAGAATTTGTCACCCGCCATTTTCATAGATTTAAGCCAA
AAGATAAATGTGTGAAAATGTACCAATGGCTGCGAAGCTTCAGGAAGTAGAGGATCCAGT
GATGCATTTTTTTTTTCTAAGGGAAAGCTGGCTCTTTAATTCAGATGCTGAATTGGTGC
CATGAAAACAGAAAATGCTATTTTCTTATTATTTAAAAGAACGTCTTATCTCATAAAATT
GACATTGTTCAAAGTCTTGTGGTGATTTGCACTATTGTTTTCTCGTATGGACCATGG

Chromosome map:
chromosome 6.

FIGURE 3D